

A Characterization of Cointegration

by

Phoebus Dhrymes, Columbia University

September 1995
released October 1995

1994-95 Discussion Paper Series No. 747

dp 94-95-747
pages 32

A CHARACTERIZATION OF COINTEGRATION

Phoebus J. Dhrymes*

Columbia University

08/07/93; revised 09/23/95

Abstract

In this paper we provide a succinct characterization of cointegration and a criterion that leads to a novel, and very simple test for cointegration which is related to the one proposed in Engle and Granger (1987) (EG).

Key words: Integrated Processes, Stationary Processes, Collinearity, Cointegration, Cointegration tests.

1 Introduction

The topic of cointegration has already developed a large, and still expanding, literature. It is certainly not our intention to recount it here, since excellent reviews, at various levels of complexity have appeared, as in Johansen (1991), Dickey, Jansen and Thornton (1992), Perman (1991), and others.

The purpose of this paper is to provide a characterization of cointegration. The major contributions of the paper are:

- i. it gives a characterization of cointegration for $I(d)$ processes;
- ii. it provides an explicit representation of the covariance matrix of $I(1)$ and $I(2)$ processes;

* This is a preliminary version and is not to be quoted, except by permission of the author. Comments, however, are welcome.

- iii. it provides the means for obtaining the covariance matrix of processes that are either fractionally integrated, or integrated of order greater than two;
- iv. it ties this topic to the work of R. A. Stone (1947), and, more importantly, avoids the excessively complicating and often opaque discussions in the literature as to what cointegration means and how to test for its presence;
- v. it suggests a novel test for cointegration based on the newly established form of the covariance matrix for $I(1)$ processes, to be more fully explored in a later paper;
- vi. it provides at least a judgemental way in which one may distinguish between $I(0)$ and $I(1)$ processes when the phenomenon noted by Stone is present.

2 Cointegrated Sequences and their Properties

2.1 Integrated Sequences

We begin with

Definition 1. Let $X = \{X_t : t \in \mathcal{N}\}$ be a stochastic sequence defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$, where \mathcal{N} is the integer lattice on \mathbb{R} i.e. the set $\{t : t = 0 \pm 1, \pm 2, \dots\}$. The sequence X is said to be **integrated of order d** , denoted by $X \sim I(d)$, if and only if¹

$$Y = \{Y_t : t \in \mathcal{N}, Y_t = (I - L)^d X_t\}, \quad (1)$$

is a (square integrable)² stationary sequence.

¹ In this, as in all subsequent definitions or discussions, it is to be understood that relations are stated in the lowest possible terms; thus in Definition 1, it is to be understood that d is the lowest possible positive number (typically integer) for which the definition is valid. It is certainly quite evident that if X is integrated of (integer) order d , it is also integrated of (integer) order d' for $d' > d$.

² The existence of second moments is not explicitly stated in discussions of integrated processes in the econometrics literature; it is, however, strongly implied by the context. We take this opportunity to note that, unless otherwise stated, stationarity will mean **strict stationarity**; when we speak of the covariance matrix of a stationary process we shall always mean that such covariance matrices are **nonsingular**, unless otherwise stated to the contrary.

The most general form we shall consider for Y_t is the so called general linear process³

$$Y_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}, \quad (2)$$

where $\{\alpha_j : j \in \mathcal{N}_0\}$, $\alpha_0 = 1$, $\sum_{j=0}^{\infty} |\alpha_j| < \infty$ is a sequence of constants, and the ϵ -sequence is a white-noise process with parameter σ^2 , to be denoted by $WN(\sigma^2)$.

The definition is precisely the same for multivariate (or vector) processes, i.e. a sequence of random elements X , as above, is $I(d)$ if and only if $Y = (I - L)^d X$ is a square integrable covariance stationary process. Again the most general case we shall deal with is

$$Y_t = \sum_{j=0}^{\infty} A_j \epsilon'_{t-j}, \quad Y_t = (I - L)^d X_t, \quad t \in \mathcal{N},$$

where the ϵ -sequence is a $MWN(\Sigma)$ process, (a multivariate white noise process with parameter Σ), and the nonstochastic matrices A_j obey, $\{A_j : j \geq 0, A_0 = I_q, \sum_{j=0}^{\infty} \|A_j\| < \infty\}$, where $\|\cdot\|$ is a suitable norm. The last condition ensures that the general linear (vector) process converges with probability one.

Noting that, for $\tau \geq 0$,

$$\Psi(\tau) = \text{Cov}(Y'_{t+\tau}, Y_t) = \sum_{j=0}^{\infty} A_{j+\tau} \Sigma A'_j,$$

$$\Psi(-\tau) = \Psi(\tau)',$$

$$\begin{aligned} \|\sum_{\tau=-\infty}^{\infty} \Psi(\tau)\| &\leq \|\Psi(0)\| + 2 \sum_{\tau=1}^{\infty} \|\Psi(\tau)\| \\ &\leq \|\Psi(0)\| + 2 \left(\sum_{j=0}^{\infty} \|A_j\| \right) \|\Sigma\| \left(\sum_{j=0}^{\infty} \|A_j\| \right), \quad (3) \end{aligned}$$

we verify that the process in question is indeed a strictly (as well as covariance) stationary process. It is also easy to see that if the process X were in existence in the indefinite past, the covariance matrix of any element, say X_t , would be undefined since it would involve, even in the

³ In this context, and as was noted in footnote one, it is to be understood that the general linear process, represented by the operator $\sum_{j=0}^{\infty} \alpha_j L^j$, or in the multivariate case $\sum_{j=0}^{\infty} A_j L^j$, **does not have a unit root factorization**, i.e., it cannot be represented by $(I - L) \sum_{j=0}^{\infty} D_j L^j$, such that $\sum_{j=0}^{\infty} \|D_j\| < \infty$.

simplest of cases with $d = 1$, the sum

$$\lim_{s \rightarrow \infty} \sum_{j=-s}^t \text{Var}(Y'_t).$$

Thus, in dealing with integrated processes we cannot operate without assuming certain initial conditions that limit the extent to which the past of the process is relevant. Typically, such initial conditions are of the form

$$X_{-j} = 0, \text{ at least for } 0 \leq j \leq d, \quad d \geq 1, \quad (4)$$

although other initial conditions can also be handled without undue complications.

The following procedure enables us to deal with any integrated process.⁴ Since we deal with

$$(I - L)^d X'_t = \eta'_t, \quad \eta'_t = \sum_{j=0}^{\infty} A_j \epsilon'_{t-j}, \quad (5)$$

we have, formally,

$$X'_t = (I - L)^{-d} \eta'_t. \quad (6)$$

The meaning of the operator $(I - L)^{-d}$ is to be understood in terms of the isomorphism between the algebra of the lag operator, L , and the algebra of polynomials in the real or complex indeterminate z . We further write formally, **for any** α ,

$$\begin{aligned} (1 - z)^\alpha &\sim 1 + (-1) \frac{\alpha}{1!} z + (-1)^2 \frac{\alpha(\alpha - 1)}{2!} z^2 + (-1)^3 \frac{\alpha(\alpha - 1)(\alpha - 2)}{3!} z^3 \\ &+ \dots + (-1)^r \frac{\alpha(\alpha - 1)(\alpha - 2) \dots (\alpha - (r - 1))}{r!} z^r. \end{aligned}$$

While it is true that this expansion does not converge for $|z| \geq 1$, the coefficients of the various powers of z are valid in the sense that if we invoke the isomorphism noted in Dhrymes (1982), the expansion of the operator $(I - L)^{-d}$ becomes, for $r = t - 1$,

$$X'_t = \sum_{s=0}^{t-1} \binom{d-1+s}{s} \eta'_{t-s}; \quad (7)$$

⁴ Note that replacing $-d$ by a context free parameter α , the argument below is valid for any α . It is, thus, a rather trivial matter to deal with fractional differencing, i.e. to examine sequences of the form $(I - L)^{-\alpha} X'_t = \eta'_t$, where α is a fraction, proper or improper.

for the special case $d = 1$ we obtain

$$X'_t = \sum_{s=0}^{t-1} \binom{s}{s} \eta'_{t-s} = \sum_{s=0}^{t-1} \eta'_{t-s} = \sum_{j=1}^t \eta'_j;$$

for $d = 2$

$$X'_t = \sum_{s=0}^{t-1} \binom{1+s}{s} \eta'_{t-s} = \sum_{s=0}^{t-1} (s+1) \eta'_{t-s} = \sum_{j=1}^t (t+1-j) \eta'_j,$$

and so on. The **same results** will be obtained if Eq. (7) is solved recursively, for $d = 1$ and $d = 2$, respectively. This is the sense in which the formal operations of Eqs. (5) and (6) are to be understood.

2.2 Cointegrated Sequences

Of special interest, in the context of integrated series, is the notion of cointegration defined below.

Definition 2. Let X be a multivariate $I(d)$ process as in Definition 1; the components of the random elements X_t are said to be **cointegrated of order d, b** , denoted by $X \sim CI(d, b)$, if and only if there exists at least one nontrivial vector β such that

$$X_t \beta = Z_t \sim I(d - b).$$

It is said to be **cointegrated of order d, b** , and of **cointegrating rank r** , denoted by $CI(d, b, r)$, if and only if there exist **exactly r linearly independent vectors**, say β_i , such that for $B = (\beta_1, \beta_2, \dots, \beta_r)$

$$X_t B = Z_t \sim I(d - b), \quad Z_t = (z_{ti}), \quad z_{ti} = X_t \beta_i, \quad i = 1, 2, 3, \dots, r. \quad (8)$$

Having defined what we wish to mean by cointegration, it would be useful to obtain a criterion by which we may determine whether a given sequence is cointegrated.

Remark 1. It is **remarkable** that although the literature on cointegration is voluminous, no characterization has been published, and no implications have been drawn, directly from its definition. It is one of the contributions of this paper that it provides such a characterization and a test based on it.

We begin with an example

Example 1. Consider the $I(1)$ sequence whose forcing function is $MMA(\infty)$, (multivariate or vector moving average of infinite order) i.e. the process $X = \{X'_t : t \in \mathcal{N}\}$ is such that

$$X'_t = X'_{t-1} + \eta'_t, \quad \eta'_t = \sum_{j=0}^{\infty} A_j \epsilon'_{t-j}, \quad A_0 = I_q, \quad (9)$$

and the ϵ -process is (multivariate white noise process with covariance Σ) $MWN(\Sigma)$; it is necessary to assume certain initial conditions, otherwise the covariance matrix of such process is, for every t , undefined. These are, essentially, that $\epsilon_{-j} = 0$, for $j \geq 0$. Given these initial conditions we find⁵

$$\text{Cov}(X'_t) = \sum_{j=0}^{t-1} (t-j) B_j = - \sum_{j=1}^{t-1} j B_j + \left(\sum_{j=0}^{t-1} B_j \right) t = \Psi_{0,t} + \Psi_{1,t} t, \quad (10)$$

where

$$S_{i,j} = \sum_{r=i}^j A_r \quad (11)$$

$$B_j = S_{0,j} \Sigma A'_j + A_j \Sigma S'_{0,j-1}, \quad j = 1, 2, \dots,$$

with the convention that $S_{0,0} = A_0 = I_q$. It may be further verified that the first and second matrices in Eq. (10) yield, respectively,

$$\begin{aligned} -\Psi_{0,t} &= \sum_{j=1}^{t-1} j B_j = S_{0,t-1} \Sigma \left(\sum_{r=0}^{t-1} r A_r \right)' + \sum_{r=1}^{t-1} S_{r,t-1} \Sigma S'_{0,r-1} \quad \text{or, alternatively,} \\ -\Psi_{0,t} &= \sum_{j=1}^{t-1} j B_j = \sum_{r=1}^{t-1} \left(S_{0,t-1} \Sigma S'_{0,t-1} - S_{0,r-1} \Sigma S'_{0,r-1} \right) \\ \Psi_{1,t} &= \sum_{j=0}^{t-1} B_j = S_{0,t-1} \Sigma S_{0,t-1}. \end{aligned} \quad (12)$$

Thus, we see that if we take the **first alternative**, and we wish the process X to be cointegrated we must have an additional condition satisfied by the forcing function of this example, viz.

$$\sum_{j=0}^{\infty} j A_j, \quad \text{must converge absolutely,}$$

⁵ Greater detail in the derivation of these results may be found in the author's *Topics in Advanced Econometrics: vol. III, Topics in Time Series*, (1995), unpublished.

i.e. that on an element by element basis⁶ $\sum_{j=0}^{\infty} j |a_{r,s}^j| < \infty$.

If we take the second option, we illustrate why specifying the forcing function as $MMA(\infty)$ (as in Engle and Granger (1987)) results in certain irritating features. For example, it is impossible to enforce the cointegrating condition for stationarity, since it is not possible with a fixed matrix B to induce the condition that

$$B' \text{Cov}(X'_t) B = B' \left(\sum_{r=0}^{t-1} S_{0,r} \Sigma S'_{0,r} \right) B$$

is independent of t , without requiring a condition that $A_s = 0$, for $s > k$! To avoid such difficulties we shall assume, in future discussion, that the forcing function is $MMA(k)$, with $k < \infty$.

Remark 2. Notice that in the example above neither Ψ_0 nor Ψ_1 are invariant with respect to t ; therefore it is really not possible, strictly speaking, to obtain cointegration. This is due to the nature of the specification of the forcing function, and the initial conditions we had imposed. If the forcing function is $MMA(k)$, however large k may be, the matrices in question will be of fixed form, for $t > k$. We shall revisit this issue below.

Returning to the case of an arbitrary integrated sequence of order d in Eq. (7), we note that the expansion therein becomes, after the change in variable $j = t - s$, the canonical representation

$$X'_t = \sum_{j=1}^t \binom{d-1+t-j}{t-j} \eta'_{j\cdot}. \quad (13)$$

and we can employ the apparatus of the previous example to obtain a representation of its covariance matrix. By definition

$$\begin{aligned} \text{Cov}(X'_t) &= \sum_{j=1}^t \sum_{j'=1}^t \binom{d-1+t-j}{t-j} \binom{d-1+t-j'}{t-j'} E \eta'_{j\cdot} \eta_{j'\cdot} \\ &= C_1(t) + C_2(t) \\ C_1(t) &= \sum_{j=1}^t \binom{d-1+t-j}{t-j}^2 E \eta'_{j\cdot} \eta_{j\cdot} \end{aligned} \quad (14)$$

⁶ Notice that this condition is ensured by convergence in norm, i.e. if

$$\lim_{N \rightarrow \infty} \sum_{j=0}^N j \|A_j\| < \infty.$$

$$C_2(t) = \sum_{\tau=1}^{t-1} \sum_{j=\tau+1}^t \gamma_{j,\tau} E(\eta'_{j-\tau} \eta_j + \eta'_j \eta_{j-\tau}) \quad (15)$$

$$\gamma_{j,\tau} = \binom{d-1+t-j+\tau}{t-j+\tau} \binom{d-1+t-j}{t-j},$$

or alternatively

$$C_1(t) = \sum_{r=0}^{t-1} \left[\sum_{j=1}^{t-r} \binom{d-1+t-j-r}{t-j-r}^2 \right] A_r \Sigma A'_r \quad (16)$$

$$\phi_{r,\tau} = \sum_{j=1}^{t-r-\tau} \binom{d-1+t+\tau-r-j}{t+\tau-r-j} \binom{d-1+t-r-j}{t-r-j} \quad (17)$$

$$C_2(t) = \sum_{\tau=1}^{t-1} \sum_{r=0}^{t-\tau} \phi_{r,\tau} (A_{r+\tau} \Sigma A'_r + A_r \Sigma A'_{r+\tau}). \quad (18)$$

A consequence of the development above is the characterization of the covariance matrix of an $I(d)$ in

Proposition 1. Let $X = \{X_t : t \in \mathcal{N}\}$ be a multivariate $I(d)$ process with suitable initial conditions. If

$$\text{Cov}(X'_t) = \Psi(t)$$

then

$$\Psi(t) = \sum_{j=0}^{2d-1} \Psi_j t^j, \quad (19)$$

where Ψ_j , $0 \leq j \leq 2d-1$, are square matrices of order q .

Proof: From Eq. (17) we have

$$\phi_{r,\tau} = \left(\frac{1}{(d-1)!} \right)^2 \sum_{j=1}^{t-r-\tau} \prod_{s=1}^{d-1} [(t-r-j+s+\tau)(t-r-j+s)], \quad (20)$$

whence it is seen that it involves the sum of a product of $d-1$ terms each of which contains a quadratic term. Hence, it involves the sum of $2(d-1)$ powers of integers. It follows therefore that $\phi_{r,\tau}$ contains at least one term which is of degree $2d-1$.

q.e.d.

2.3 Characterization of Cointegration

In the preceding sections we gave a definition of cointegration; the definition is typically a simple, as well as an easily grasped and extended statement of the concept we wish to define. In the contemporary usage of mathematics, the definition is quickly followed by an operational criterion which gives succinctly a relatively simple procedure (criterion) by which one may decide whether a given entity conforms to the requirements of the definition. For example one may define the (column/row) rank of a matrix as the number of **linearly independent** (columns/rows) it contains. This is a simple and easily grasped notion, but does not immediately give rise to a relatively simple procedure for determining the rank of a matrix. In the case of **square matrices** (of order n), we have the characterization that such a matrix is of rank n if and only if its determinant is nonnull.

Such a characterization is absent from the econometric literature of cointegration, where much of the published work is devoted to establishing the validity of certain **implications** of cointegration. Our task in this section is to establish such a characterization. We have

Proposition 2. Let X be a (multivariate) $I(d)$ process as in Definition 1; X is $CI(d, d, r)$ if and only if⁷

$$\text{Cov}(X'_t) = \Psi(t) = \Psi_0 + \Psi_1(t), \quad (21)$$

such that

- i. Ψ_0 does not depend on t , $\text{rank}(\Psi_0) \geq r$ and the intersection of the null space of Ψ_0 and $\Psi_1(t)$ consists only of the zero vector;
- ii. $\text{rank}(\Psi_1(t)) = q - r$, i.e. there exists an appropriately dimensioned matrix B of rank, at most, r such that $\Psi_1(t)B = 0$.

Proof: Since for **fixed** t , given the initial conditions imposed in the econometric literature, the covariance matrix of X_t is bounded, while it becomes unbounded with t , we may without loss of generality write

$$\Psi(t) = \Psi_0 + \Psi_1(t).$$

Necessity: if X is $CI(d, d, r)$, there exists a matrix B of rank r such that for $Z_t = X_t B$,

$$\text{Cov}(Z'_t) = B' \Psi_0 B + B' \Psi_1(t) B = B' \Psi_0 B.$$

⁷ These conditions are to be understood in the context of Remark 2.

The expression above defines Z_t as a square integrable stationary process,⁸ only if $B'\Psi_1(t)B = 0$, and $B'\Psi_0B > 0$. Thus, $\Psi_1(t)$ is of rank $q - r$ and, since B spans the null space of $\Psi_1(t)$, the intersection of the latter and the null space of Ψ_0 consists only of the null vector. Moreover, $\text{rank}(\Psi_0) \geq r$, otherwise $B'\Psi_0B$ need not be positive definite, which completes the proof of necessity.

Sufficiency: suppose conditions i and ii are satisfied; then, there exists a suitably dimensioned matrix, B of rank, at most, r such that $Z_t = X_t B$ is a stationary process with fixed covariance matrix $B'\Psi_0B > 0$, or equivalently $Z_t \sim I(0)$.

q.e.d.

Remark 3. If one can prove the converse of Proposition 2, viz. that **if the covariance matrix of a process is a (matrix) polynomial of degree $2d - 1$, then the process in question is $I(d)$** , Proposition 2, may be modified into a characterization of $CI(d, b, r)$. Otherwise we are left with

Corollary 1. Let X be an $I(d)$ process, as in Proposition 2, and suppose there exists a suitably dimensioned matrix, B , of rank r such that $Z_t = X_t B$ is $I(d - b)$, i.e. X is $CI(d, b, r)$; then the covariance matrix of the cointegral⁹ vector Z_t is a matrix polynomial of degree $2(d - b) - 1$.

Proof: Obvious from Propositions 1 and 2.

q.e.d.

Example 2. Consider the (bivariate) model given in Engle and Granger (1987)

$$X_t \begin{bmatrix} 1 & 1 \\ \beta & \alpha \end{bmatrix} = u_t, \quad (I - L)u_{t1} = \epsilon_{t1}, \quad u_{t2} = [I/(I - \rho L)]\epsilon_{t2}.$$

Manipulating the expression above we find

$$(I - L)X_t \begin{bmatrix} 1 & 1 \\ \beta & \alpha \end{bmatrix} = (\epsilon_{t1} \quad \frac{I - L}{I - \rho L} \epsilon_{t2});$$

⁸ This is another instance in which the cointegration literature in econometrics is ambiguous. In our characterization we require that $B'\Psi_0B > 0$; otherwise cointegration becomes **precisely** collinearity.

⁹ In the literature, there does not appear that there is a specific term to describe the vector Z_t . In the spirit of this topic I have come to call it the *cointegral* vector, as distinct from the *cointegrated* vector X_t .

Further simplification yields

$$(I - L)X_t' = C(L)\epsilon_t, \quad C(L) = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha I & -\frac{\beta(I-L)}{I-\rho L} \\ -I & \frac{I-L}{I-\rho L} \end{bmatrix},$$

$$C(I) = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha I & 0 \\ -I & 0 \end{bmatrix}, \quad C^*(L) = \frac{1}{\alpha - \beta} \begin{bmatrix} 0 & \frac{\beta I}{I-\rho L} \\ 0 & \frac{I}{I-\rho L} \end{bmatrix},$$

$$C(L) = C(I) + (I - L)C^*(L).$$

Amplifying the statement in Remark 2, we note that the covariance matrix of the process above is not of the form $\Psi_t = \Psi_{(0)} + \Psi_{(1)}t$, with Ψ_0 and Ψ_1 , **independent of t** . To see this note that **imposing the initial condition** $X_0 = 0$, which implies $\epsilon_{-i} = 0$ for $i \geq 0$, we find

$$X_t = \frac{1}{(\alpha - \beta)} (\epsilon_{t1}^*, \epsilon_{t2}^*) \begin{bmatrix} \alpha & -1 \\ -\beta & 1 \end{bmatrix}, \quad (22)$$

where $\epsilon_{t1}^* = \sum_{s=1}^t \epsilon_{s1}$ and $\epsilon_{t2}^* = \sum_{j=0}^{t-1} \rho^j \epsilon_{t-j,2}$. It is now easy to establish

$$E\epsilon_{t1}^{*2} = t\phi_{11}, \quad \phi_{ij} = E\epsilon_{ti}\epsilon_{tj}, \quad i, j = 1, 2;$$

$$E\epsilon_{t2}^{*2} = \phi_{22}\rho_2, \quad \rho_2 = \sum_{j=0}^{t-1} \rho^{2j}, \quad \rho_1 = \sum_{j=0}^{t-1} \rho^j$$

$$E\epsilon_{t1}^* \epsilon_{t2}^* = \phi_{12}\rho_1$$

and that

$$\text{Cov}(X_t') = \frac{1}{(\alpha - \beta)^2} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \text{Cov} \begin{pmatrix} \epsilon_{t1}^* \\ \epsilon_{t2}^* \end{pmatrix} \begin{bmatrix} \alpha & -1 \\ -\beta & 1 \end{bmatrix}$$

$$= \Psi_{(0)} + \Psi_{(1)}t,$$

where

$$\Psi_{(0)} = \frac{1}{(\alpha - \beta)^2} \begin{bmatrix} \beta^2\gamma_2 - 2\alpha\beta\gamma_1 & (\alpha + \beta)\gamma_1 - \beta\gamma_2 \\ (\alpha + \beta)\gamma_1 - \beta\gamma_2 & \gamma_2 - 2\gamma_1 \end{bmatrix},$$

$$\gamma_1 = \phi_{12}\rho_1, \quad \gamma_2 = \phi_{22}\rho_2,$$

$$\Psi_{(1)} = \frac{1}{(\alpha - \beta)^2} \begin{bmatrix} \alpha^2\phi_{11} & -\alpha\phi_{11} \\ -\alpha\phi_{11} & \phi_{11} \end{bmatrix}. \quad (23)$$

It is clear that $\Psi_{(1)}$ is, for each t , of rank 1 and hence **singular**. The characteristic vector corresponding to the zero root is $(1, \alpha)'$. Consequently, the cointegral scalar $z_t = X_t(1, \alpha)'$ obeys

$$Ez_t = 0, \quad \text{for all } t, \quad \text{Var}(z_t) = (1, \alpha)\Psi_{(0)}(1, \alpha)' = \gamma_2,$$

so that **it does not have a stationary distribution since its variance depends on t** , through the entity ρ_2 . It is evident, however, that this is due solely to the imposition of initial conditions, and the fact that the forcing function for the second equation is $MA(\infty)$, of a special form, corresponding to the inverse, or final form representation of an $AR(1)$ process; it also evident that

$$\lim_{t \rightarrow \infty} \rho_2 = \frac{1}{1 - \rho^2}, \quad \lim_{t \rightarrow \infty} \gamma_2 = \frac{\phi_{22}}{1 - \rho^2},$$

and thus

$$\lim_{t \rightarrow \infty} \text{Var}(z_t) = \frac{\phi_{22}}{1 - \rho^2}, \quad (24)$$

as was to be expected by construction.

Remark 4. As the preceding example makes clear, it is not only convenient but necessary as well, to fix the extent of the moving average in the forcing function, i.e. to specify a forcing function which is $MMA(k)$. This, will allow all standard definitions of cointegration to stand without the complicating features, noted in Remark 2 and illustrated in Example 2. Incidentally note that, strictly speaking, in the exposition by EG (1987) it is not possible to have cointegration, as the term is currently defined.

Example 3. In light of Remark 4, consider the process $X = \{X_t : t \in \mathcal{N}\}$, given by

$$Y_t = (I - L)^2 X_t, \quad Y_t = \eta_t, \quad \eta'_t = \sum_{j=0}^k A_j \epsilon'_{t-j}.$$

so that we have the representation

$$X'_t = 2X'_{t-1} - X'_{t-2} + \sum_{j=0}^k A_j \epsilon'_{t-j}, \quad A_0 = I_q, \quad \text{Cov}(\epsilon'_t) = \Sigma,$$

which identifies X as an $I(2)$ process. Moreover, if we assume the initial conditions $X_0 = X_{-1} = 0$, we may determine the behavior of the covariance matrix of this $I(2)$ sequence by the same method as above.

Solving the equation above, we obtain the canonical representation

$$X'_t = \sum_{j=1}^t (t+1-j) \eta_j, \quad (25)$$

whence we conclude

$$\text{Cov}(X'_t) = C_1(t) + C_2(t) \quad (26)$$

$$C_1(t) = \sum_{j=1}^t (t+1-j)^2 E\eta'_j \eta_j.$$

$$C_2(t) = \sum_{\tau=1}^k \sum_{j=1}^{t-\tau} [(t+1-j)^2 + \tau(t+1-j)] E(\eta_{j-\tau} \eta_j + \eta'_j \eta_{j-\tau}).$$

Moreover, it may further be shown that

$$C_1(t) = \sum_{r=0}^k \left(\sum_{j=1}^{t-r} (t+1-r-j)^2 A_r \Sigma A'_r \right) \quad (27)$$

$$\phi_{r,\tau} = \sum_{j=1}^{t-\tau-r} [(t+1-r-\tau-j)^2 + \tau(t+1-r-\tau-j)],$$

$$C_2(t) = \sum_{\tau=1}^k \left[\sum_{r=0}^{k-\tau} (\phi_{r,\tau} (A_{r+\tau} \Sigma A'_r + A_r \Sigma A'_{r+\tau})) \right], \quad (28)$$

which completes the derivation of the covariance matrix of an element of an $I(2)$ process with forcing function which is $MM A(k)$.

While the formal derivation is now complete, the representation above is not particularly informative. To gain some insight into its structure we note that

$$\phi_{r,\tau} = \alpha_{r+\tau} - \tau \beta_{r+\tau} \quad (29)$$

$$\alpha_{r+\tau} = \frac{(t-r-\tau)(t-r-\tau+1)(2t-2r-2\tau+1)}{6},$$

$$\beta_{r+\tau} = \frac{(t-r-\tau)(t-r-\tau+1)}{2},$$

$$\alpha_r = \frac{(t-r)(t-r+1)(2t-2r+1)}{6},$$

and, consequently, that

$$C_1(t) = \sum_{r=0}^k \alpha_r A_r \Sigma A'_r, \quad (30)$$

$$C_2(t) = \sum_{\tau=1}^k \sum_{r=0}^{k-\tau} \alpha_{r+\tau} (A_{r+\tau} \Sigma A'_r + A_r \Sigma A'_{r+\tau})$$

$$+ \sum_{\tau=1}^k \tau \left(\sum_{r=0}^{k-\tau} \beta_{r+\tau} (A_{r+\tau} \Sigma A'_r + A_r \Sigma A'_{r+\tau}) \right). \quad (31)$$

$$(32)$$

Making the change in variable $j = r + \tau$, $\tau = \tau$, we note that the range of the two indices is given by $1 \leq j \leq k$, $1 \leq \tau \leq j$, due to the fact that $r = j - \tau$. In this notation we obtain the representation

$$C_1(t) = \sum_{r=0}^k \alpha_r A_r \Sigma A'_r \quad (33)$$

$$C_2(t) = \sum_{j=1}^k \alpha_j (A_j \Sigma S'_{0,j-1} + S_{0,j-1} \Sigma A'_j) + \sum_{j=1}^k \beta_j (A_j \Sigma P'_{j-1} + P_{j-1} \Sigma A'_j)$$

$$S_{i,j} = \sum_{s=i}^j A_s, \quad P_{j-1} = \sum_{s=0}^{j-1} (j-s) A_s, \quad \text{so that}$$

$$C(t) = \sum_{j=0}^k \alpha_j B_j + \sum_{j=1}^k \beta_j (A_j \Sigma P'_{j-1} + P_{j-1} \Sigma A'_j). \quad (34)$$

Remark 5. In much of the econometrics literature the concept of cointegration is linked to the “long run equilibrium relationship” among economic (mostly macro) variables. While this could possibly be an explanation as to **why** certain variables may be cointegrated, the concept of cointegration is basically a mathematical, more precisely a probabilistic one. In attempting to elucidate the implication of this relationship it is **the mathematical-probabilistic** aspects of the definition that are paramount.

To form some intuition regarding the concept conveyed by cointegration, it is useful to compare it with **collinearity**, or **linear dependency**. If we rank the “randomness” of an entity by the magnitude of its variability, say its variance, a zero mean random vector, say z , with nonsingular covariance matrix is not “degraded” by a **nonsingular linear transformation**. For example, if A is nonsingular, Az still has a nonsingular covariance matrix. On the other hand, **if the covariance matrix is singular** there exists at least one vector, say β such that $\text{Var}(\beta' z) = 0$; when this is so, the elements of z are said to be **collinear**, or to exhibit **linear dependencies**, so that the degree of “randomness” of z has been “degraded”. The concept of cointegration conveys a similar notion. For example, if X is $CI(2, 1)$, of (cointegrating) rank 1, it means that the components of the vector X_t are each $I(2)$, but there exists a nontrivial vector, say β , which degrades the randomness of X_t to $I(1)$, since

$X_t \beta \sim I(1)$. Similarly if X is $CI(2, 2)$ of (cointegrating) rank 1, there exists a vector, say γ such that $X_t \gamma \sim I(0)$, i.e. the linear combination in question is a square integrable stationary process.

Example 4. Consider the $I(1)$ process X of Example 1, with a forcing function which is $MMA(k)$, instead of $MMA(\infty)$, i.e. such that $\eta'_t = \sum_{j=0}^k A_j \epsilon'_{t-j}$. Its covariance matrix, for $t \geq k+1$, is given by

$$\text{Cov}(X'_t) = \Psi_0 + \Psi_1 t, \quad -\Psi_0 = \sum_{s=1}^k s B_s, \quad \Psi_1 = \sum_{s=0}^k B_s.$$

Moreover, it is easily shown that

$$\sum_{j=1}^k j B_j = \sum_{r=1}^k \left(S_{0,k} \Sigma S'_{0,k} - S_{0,r-1} \Sigma S'_{0,r-1} \right) = k S_{0,k} \Sigma S'_{0,k} - \sum_{r=1}^k S_{0,r-1} \Sigma S'_{0,r-1}. \quad (35)$$

Consequently, we have

$$\text{Cov}(X'_t) = (t - k) S_{0,k} \Sigma S'_{0,k} + \sum_{r=1}^k S_{0,r-1} \Sigma S'_{0,r-1}. \quad (36)$$

Three important implications arise from the representation in Eq. (36). First, for the $I(1)$ process above to be $CI(1, 1, r)$ there must exist a suitably dimensioned matrix B such that $S'_{0,k} B = 0$ **and** $S'_{0,j} B \neq 0$ for at least one $j = 0, 1, 3, \dots, k-1$; the last condition is automatically satisfied since it is assumed as a matter of convention and identification that $A_0 = I_q$.

Second, the process X **cannot possibly be cointegrated** if $k = 0$.

Third, the covariance stationarity of the cointegral vector is easily demonstrated, by noting that the cointegral vector may be represented¹⁰ as

$$Z'_t = B' \xi'_t, \quad \text{where} \quad (37)$$

$$\xi'_t = \sum_{r=0}^{k-1} S_{0,r} \zeta'_{t-r}, \quad \text{or} \quad \sum_{r=1}^k S_{r,k} \zeta^{*'}_{t-r}.$$

$$\begin{aligned} \text{Cov}(Z'_{t+\tau}, Z'_t) &= B' \left(\sum_{r=0}^{k-1} \sum_{r'=0}^{k-1} S_{0,r} \left[E \zeta'_{t+\tau-r} \zeta'_{t-r'} \right] S'_{0,r'} \right) B \\ &= B' \left(\sum_{r=\tau}^{k-1} S_{0,r} \Sigma S'_{0,r-\tau} \right) B, \end{aligned}$$

¹⁰ In the representation below ζ or ζ^* is a $MWN(\Sigma)$.

which validates the claim that the cointegral process is indeed **(covariance) stationary**.

Remark 6. Before we leave this topic we shall (i) clarify the role played by initial conditions, in the case of an $I(1)$ process whose forcing function is $MMA(\infty)$; (ii) establish the nature of the argument in passing to the limit for the case where the forcing function is $MMA(\infty)$ vis-avis the case where it is $MMA(k)$; and, finally, (iii) we shall delineate the similarities and the differences between the two.

To this end, consider $X'_t = \sum_{j=1}^t \eta'_{j\cdot}$, $\eta'_{j\cdot} = \sum_{s=0}^{\infty} A_s \epsilon'_{j-s\cdot}$. In contrast to the $MMA(k)$, **no matrix A_s is known a priori to be zero**. As before

$$\text{Cov}(X'_t) = \sum_{j=1}^t E\eta'_{j\cdot}\eta_{j\cdot} + \sum_{j \neq j'} E\eta'_{j\cdot}\eta_{j'\cdot} = C_1(t) + C_2(t).$$

We note that

$$\begin{aligned} C_1(t) &= \sum_{j=1}^t \sum_{s=0}^{\infty} \sum_{s'=0}^{\infty} A_s \left(E\epsilon'_{j-s\cdot}\epsilon_{j-s'\cdot} \right) A'_{s'}, \text{ nonnull only for } s = s' \\ &= \sum_{j=1}^t \sum_{s=0}^{j-1} A_s \left(E\epsilon'_{j-s\cdot}\epsilon_{j-s\cdot} \right) A'_s, \text{ and changing the order of summation} \\ &= \sum_{s=0}^{t-1} \sum_{j=s+1}^t A_s \left(E\epsilon'_{j-s\cdot}\epsilon_{j-s\cdot} \right) A'_s = \sum_{s=0}^{t-1} (t-s) A_s \Sigma A'_s. \end{aligned}$$

Next, we note that

$$C_2(t) = \sum_{j=2}^t \sum_{\tau=1}^{j-1} E \left(\eta'_{j-\tau\cdot}\eta_{j\cdot} + \eta'_{j\cdot}\eta_{j-\tau\cdot} \right) = C_{21}(t) + C_{22}(t);$$

Since $C_{22}(t) = C_{21}(t)'$ we need only evaluate $C_{21}(t)$. Thus,

$$\begin{aligned} C_{21}(t) &= \sum_{j=2}^t \sum_{\tau=1}^{j-1} \sum_{s=0}^{\infty} \sum_{s'=0}^{\infty} A_s \left(E\epsilon'_{j-\tau-s\cdot}\epsilon_{j-s'\cdot} \right) A'_{s'}, \text{ nonnull only for } s' = s + \tau \\ &= \sum_{j=2}^t \sum_{\tau=1}^{j-1} \sum_{s=0}^{\infty} A_s \left(E\epsilon'_{j-\tau-s\cdot}\epsilon_{j-\tau-s\cdot} \right) A'_{s+\tau}, \text{ put } \tau + s = i, \tau = \tau, \\ &= \sum_{j=2}^t \sum_{i=1}^{j-1} \sum_{\tau=1}^i A_{i-\tau} \left(E\epsilon'_{j-i\cdot}\epsilon_{j-i\cdot} \right) A'_i, \text{ change order of summation} \\ &= \sum_{i=1}^{t-1} S_{0,i-1} \sum_{j=i+1}^t \left(E\epsilon'_{j-i\cdot}\epsilon_{j-i\cdot} \right) A'_i = \sum_{i=1}^{t-1} (t-i) S_{0,i-1} \Sigma A'_i. \end{aligned}$$

It follows, therefore, that

$$C(t) = C_1(t) + C_2(t) = \sum_{j=0}^{t-1} (t-j) B_j = \Psi_{0,t} + \Psi_{1,t},$$

where, now,

$$-\Psi_{0,t} = \sum_{j=0}^{t-1} j B_j, \quad \Psi_{1,t} = \sum_{j=0}^{t-1} B_j.$$

Comparing to Eqs. (35) and (36), we see that they are basically of the same form, except that the upper limit of the sums is not truncated at k . Thus, letting $k \rightarrow \infty$ cannot be indulged in as **an independent activity**. This is due to the nature of the initial conditions we have imposed. We also have the representation

$$\Psi_{1,t} = S_{0,t-1} \Sigma S'_{0,t-1}, \quad -\Psi_{0,t} = S_{t-1} \Sigma \left(\sum_{r=1}^{t-1} r A_r \right)' + \sum_{r=1}^{t-1} S_{r,t-1} \Sigma S'_{0,r-1}$$

or, alternatively,

$$-\Psi_{0,t} = \sum_{r=1}^{t-1} \left(S_{0,t-1} \Sigma S'_{0,t-1} - S_{0,r-1} \Sigma S'_{0,r-1} \right).$$

If we take the first alternative in the representation of $\Psi_{0,t}$, we need to invoke the condition just after Eq. (12), or that in footnote 6, in order to ensure that the cointegral vector is well defined.¹¹ This condition was also noted, in the EG context, by Stock (1987) and Phillips and Solo (1992). If we take the second alternative, the covariance matrix becomes

$$\text{Cov}(X'_t) = S_{0,t-1} \Sigma S'_{0,t-1} + \sum_{r=1}^{t-1} S_{0,r-1} \Sigma S'_{0,r-1};$$

if we assume that the first matrix converges absolutely and, moreover, that

$$B' S_{0,\infty} \Sigma S'_{0,\infty} B = 0,$$

which is the natural extension of cointegration requirement when the forcing function is $MMA(k)$, then the second term of the covariance matrix above must become unbounded as $t \rightarrow \infty$. But this would mean that we could not confidently assert that the cointegral vector has a bounded, i.e. asymptotically finite, covariance matrix.

¹¹ Even though this condition resolves the problem for the cointegral vector, the first term of the covariance matrix remains quite problematic.

Thus, it would appear that in most of the literature, strictly speaking, all results are illusory, since they are carried out on the basis of a forcing function which is $MMA(\infty)$. In effect, in EG the underlying model is (implicitly)

$$(I - L)X'_t = [\Pi(L)]^{-1}A(L)\epsilon'_t, \quad \Pi(L) = \sum_{j=0}^p \Pi_j L^j, \quad A(L) = \sum_{j=0}^k A_j,$$

where $A(L)$ is a **noninvertible operator**. In Johansen, the model is

$$\Pi(L)X'_t = \epsilon'_t,$$

and it is clear that, given the standard initial conditions, the covariance matrix of the cointegral vector cannot, in **either case**, have a stationary form, as is illustrated in Example 2. Evidently, we could avoid all problems noted above if we amended the definition of cointegration so that it **requires the cointegral vector only to have a bounded (not a stationary) covariance matrix**. This, however, may well create problems currently unanticipated.

Example 5. If, in Example 4, we take the process to be $I(2)$ and the forcing function to be $MMA(k)$, the covariance matrix is given by

$$\text{Cov}(X'_t) = \sum_{j=0}^k \alpha_j B_j + \sum_{j=1}^k \beta_j (A_j \Sigma P'_{j-1} + P_{j-1} \Sigma A'_j). \quad (38)$$

Since

$$B_j = S_{0,j} \Sigma S_{0,j} - (S_{0,j-1} \Sigma S'_{0,j} - S_{0,j-1} \Sigma A'_j) = S_{0,j} \Sigma S'_{0,j} - S_{0,j-1} \Sigma S'_{0,j-1}, \quad (39)$$

and

$$\alpha_j - \alpha_{j+1} = (t - j)^2, \quad 0 \leq j \leq k - 1, \quad (40)$$

we may rewrite the covariance matrix of Eq. (38) as

$$\begin{aligned} \text{Cov}(X'_t) = & [\alpha_k - (t - k)^2] S_{0,k} \Sigma S'_{0,k} + \sum_{j=0}^{k-1} (t - j)^2 S_{0,j} \Sigma S'_{0,j} \\ & + \sum_{j=0}^k (t - j)^2 \left(S_{0,j} \Sigma S'_{0,j} + \frac{A_j \Sigma P'_{j-1} + P_{j-1} \Sigma A'_j}{2} \right) \\ & + \sum_{j=1}^k (t - j) \frac{A_j \Sigma P'_{j-1} + P_{j-1} \Sigma A'_j}{2}, \end{aligned} \quad (41)$$

it being understood that $P_{j-1} = 0$, for $j < 1$.

We conclude this example by giving a representation of the covariance matrix of $I(2)$ processes with forcing function $MM A(k)$, for $k = 0, 1, 2$. For $k = 0$ we have

$$\text{Cov}(X'_t) = \alpha_0 A_0 \Sigma A_0 = \frac{t(t+1)(2t+1)}{6} \Sigma, \quad (42)$$

and hence there is **no cointegration except in the form of collinearity**.

For $k = 1$ we find

$$\begin{aligned} \text{Cov}(X'_t) &= \alpha_0 B_0 + \alpha_1 B_1 + \beta_1 (A_1 \Sigma P'_0 + P_0 \Sigma A'_1) \\ &= \frac{t(t+1)(t-1)}{3} S_{0,1} \Sigma S'_{0,1} + \frac{t(t+1)}{2} A_0 \Sigma A'_0 - \frac{t(t-1)}{2} A_1 \Sigma A'_1, \end{aligned} \quad (43)$$

and again there is no cointegration except in the form of collinearity.

For $k = 2$ we obtain from Eq. (41)

$$\begin{aligned} \text{Cov}(X'_t) &= \alpha_0 B_0 + \alpha_1 B_1 + \alpha_2 B_2 + \beta_1 (A_1 \Sigma A'_0 + A_0 \Sigma A'_1) \\ &\quad + \beta_2 (A_1 \Sigma A'_2 + A_2 \Sigma A'_1 + 2A_0 \Sigma A'_2 + 2A_2 \Sigma A'_0). \end{aligned} \quad (44)$$

To facilitate the simplification of the equation above we note that

$$A_1 \Sigma A'_0 + A_0 \Sigma A'_1 = B_1 - A_1 \Sigma A'_1 \quad (45)$$

and moreover that the **coefficient of** β_2 is given by

$$H = S_{0,2} \Sigma S'_{0,2} - S_{0,1} \Sigma S'_{0,1} + (A_0 + A_1) \Sigma (A_0 + A_1)' - 2A_2 \Sigma A'_2 - A_0 \Sigma A'_0. \quad (46)$$

Noting that $B_2 = S_{0,2} \Sigma S'_{0,2} - S_{0,1} \Sigma S'_{0,1}$, the covariance matrix in Eq. (43) may be written as

$$\begin{aligned} \text{Cov}(X'_t) &= \alpha_0 B_0 + \alpha_1 B_1 + (\alpha_2 + \beta_2) B_2 + \beta_1 (B_1 - A_1 \Sigma A'_1) \\ &\quad + \beta_2 (A_0 + A_2) \Sigma (A_0 + A_1)' - 2\beta_2 A_2 \Sigma A'_2 - \beta_2 A_0 \Sigma A'_0 \\ &= (\alpha_2 + \beta_2) S_{0,2} \Sigma S'_{0,2} + [(\alpha_1 + \beta_1) - (\alpha_2 + \beta_2)] S_{0,1} \Sigma S'_{0,1} \\ &\quad + \beta_2 (A_0 + A_2) \Sigma (A_0 + A_2)' \\ &\quad - 2\beta_2 A_2 \Sigma A'_2 - \beta_1 A_1 \Sigma A'_1 + (\alpha_0 - \alpha_1 - \beta_1 - \beta_2) A_0 \Sigma A'_0. \end{aligned} \quad (47)$$

Evaluating the coefficients we obtain

$$\begin{aligned} \text{Cov}(X'_t) &= \frac{t(t-1)(t-2)}{3} S_{0,2} \Sigma S'_{0,2} + t(t-1) S_{0,1} \Sigma S'_{0,1} \\ &\quad + \frac{(t-1)(t-2)}{2} (A_0 + A_2) \Sigma (A_0 + A_2)' \\ &\quad - (t-1)(t-2) A_2 \Sigma A'_2 - \frac{t(t-1)}{2} A_1 \Sigma A'_1 + (2t-1) A_0 \Sigma A'_0, \end{aligned} \quad (48)$$

which may also be expressed in the more revealing form

$$\begin{aligned} \text{Cov}(X'_t) = & \frac{t(t-1)(t-2)}{3} S_{0,2} \Sigma S'_{0,2} + t(t-1) \left(S_{0,1} \Sigma S'_{0,1} - \frac{A_1 \Sigma A'_1}{2} \right) \\ & + (t-1)(t-2) \left(\frac{(A_0 + A_2) \Sigma (A_0 + A_2)'}{2} - A_2 \Sigma A'_2 \right) \\ & + (2t-1) A_0 \Sigma A'_0. \end{aligned} \quad (49)$$

Two observations are worthwhile. First, it is possible for the process to be $CI(2, 1, r)$ but **not** $CI(2, 2, r)$, except in the case of **collinearity**; this is so since there is no distinct constant term, and the last term is $(2t-1)\Sigma$, due to the fact that $A_0 = I_q$. Thus $CI(2, 2, r)$ is possible **only through collinearity**. The second observation is to note the really special conditions required for $CI(2, 1, r)$, even in the case $k = 2$. In particular, it is required that the cointegrating matrix, say B , be in the null space of $S_{0,2}$ and contain the common characteristic vectors of the (four) matrices in round brackets; moreover the corresponding characteristic roots of the two pairs of matrices must be multiples of each other, by a factor of two!

2.4 Previous Empirical Implementations

EG essentially recommend an implementation of the preceding through the covariance matrix

$$M_T = \frac{1}{T^2} X' X, \quad X = (X_t), \quad t = 1, 2, \dots, T. \quad (50)$$

This suggestion was later amplified and modified in Watson (1987), Stock and Watson (1989), and ultimately by Johansen (1988), Johansen and Juselius (1990), and Johansen (1992), (J), whose approach takes full advantage of the assumption implicit in EG, that $C(L)$, in the EG notation, is of the form $[\Pi(L)]^{-1}$, where $\Pi(L)$ is a matrix polynomial of degree p in the lag operator L . J 's approach leads to canonical variates or reduced rank regression methods. EG originally suggested that the cointegrating vectors be obtained as the characteristic vectors corresponding to the r smallest (or zero?) roots of M_T but they had no convincing justification for that. In Stock (1987) we have a clearer explication of the processes, while in Stock and Watson (1989) we have a more complex and rather opaque justification for the estimation (and test) of cointegration (vectors), involving "common trends". Finally, in Johansen we place the problem in the context of a vector autoregressive model (of order p) to begin with (which is not subject to test), and in the **context**

of this maintained hypothesis we examine whether the residuals in the regression of $\Delta X'_t$ and X'_{t-p} on $\Delta X_{t-1}, \Delta X_{t-2}, \dots, \Delta X_{t-p+1}$, yield a reduced set of canonical variates. The purpose of examining these two sets of residuals is to determine how many canonical variates can be defined, i.e. how many characteristic roots in a certain context can be asserted to be zero. If none can be so asserted, we may define q pairs of canonical variates, and thus we have **no cointegration**; if some roots, say r , may be asserted to be null, **we have cointegration of rank r** . Alternatively, we are asking whether in a (random) linear representation of the two sets of residuals the matrix connecting them is of full or reduced rank. If the rank is **maximally q** but the actual rank is $q - r$, **we have cointegration of rank r** . If the rank in question is the maximal possible, **we do not have cointegration**. The connection between reduced rank regressions and the problem of canonical variates was first noted by Anderson (1951), (1976), Izenman (1975), and Tso (1981). Notice further that acceptance of cointegration, in this context, need not imply what is commonly understood by the term, viz. that having dealt with a number of $I(1)$ variables we have now discovered precisely those linear combinations that represent nontrivial relationships among them, and which are $I(0)$. Indeed, in this context, the $I(1)$ character of the series is **given a priori** and the only question the test settles is whether such series are **cointegrated**. In practice, however, such results, far from establishing cointegration in the generally understood sense, may simply be a reflection of those found by Stone (1947), and discussed in the author's early work Dhrymes (1970).

3 A New Test for Cointegration

The test suggested by the discussion in this paper, rests on the development in sections 2.2 and 2.3. As in the previous literature, we shall primarily derive a test for cointegration in $I(1)$ processes, and merely indicate the analogous procedure for $I(2)$ and higher order processes.

It is certainly not our intention here to compare intensively the estimators proposed by EG (1987), Stock (1987), Stock and Watson (1989), or Johansen (1988), (1991), Johansen and Juselius (1990), to mention but a few, to the estimator implied by the preceding discussion. Here we shall merely formulate the test and indicate, in broad outline, its distributional characteristics; in a subsequent paper, we shall investigate more thoroughly the implied test, the distribution of the relevant test statistic, and compare it in terms of these aspects to the set of tests just noted.

We begin by noting that, from previous discussions, we had estab-

lished that if we specify the forcing function to be $MMA(k)$ ¹² we have

$$\begin{aligned}
\text{Cov}(X'_t) &= \sum_{r=0}^{t-1} S_{0,r} \Sigma S'_{0,r}, \quad \text{if } t \leq k \\
&= (t-k) S_{0,k} \Sigma S'_{0,k} + \sum_{r=0}^{k-1} S_{0,r} \Sigma S'_{0,r}, \quad \text{if } t \geq k+1, \text{ or} \\
&= t(S_{0,k} \Sigma S'_{0,k}) + D, \quad D = \sum_{r=0}^{k-1} (S_{0,k} \Sigma S'_{0,k} - S_{0,r} \Sigma S'_{0,r}).
\end{aligned} \tag{51}$$

Our objective is to test whether the matrix $S_{0,k} \Sigma S'_{0,k}$ is nonsingular; if it is, there is **no cointegration**; if it is not, the components of the vector X_t **are cointegrated**, with cointegrating rank say r , and the cointegrating matrix is the matrix of **characteristic vectors** corresponding to the r **smallest (zero) characteristic roots**. Unfortunately, such cointegrating vectors are **not unique** subject to a normalization, as is usually the case with characteristic vectors. If the reader requires a rationale for this result, set up the following problem:

$$\text{Minimize } \beta' (S_{0,k} \Sigma S'_{0,k}) \beta, \quad \text{subject to } \beta' \beta = 1. \tag{52}$$

Since the minimand is bounded below by zero, the global minimum is attained if we can find a vector that attains this value, which leads us to consider the equation

$$|\lambda I_q - S_{0,k} \Sigma S_{0,k}| = 0.$$

Evidently, if $\lambda_1, \beta_{.1}$ is a pair of characteristic root and associated characteristic vector, they obey

$$\beta'_{.1} S_{0,k} \Sigma S_{0,k} \beta_{.1} = \lambda_1.$$

Consequently, we choose $\lambda_1 = 0$. We may do so again trying to determine another vector β that minimizes the expression in Eq. (60), subject to the same normalization, **and** $\beta' \beta_{.1} = 0$. Alternatively, we may set up the general problem as follows:

$$\text{Minimize } \text{tr} B' (S_{0,k} \Sigma S'_{0,k}) B, \quad \text{subject to } B' B = I_r. \tag{53}$$

If we write the characteristic roots in terms of increasing order of magnitude, the solution to this problem entails the characteristic vectors corresponding to the r smallest roots and

$$B' (S_{0,k} \Sigma S'_{0,k}) B = \Lambda_{(r)}, \quad \Lambda_{(r)} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r).$$

¹² Note that in this framework, k need not be known, since what matters is that it is finite and **not its numerical magnitude**.

If $\lambda_r = 0$, the value of the minimand is zero, but it may very well be that $\lambda_{r+1} = 0$ as well, so that it becomes clear that the problem has not been defined very well. Consequently, the formulation becomes

Find a matrix, B , of maximal rank such that it minimizes

$$\text{tr} B'(S_{0,k} \Sigma S'_{0,k}) B, \text{ subject to } B' B = I. \quad (54)$$

The test suggested by this procedure is the following: Find an (at least asymptotically unbiased) estimator of $S_{0,k} \Sigma S'_{0,k}$; find the **maximal number** of zero characteristic roots, and their associated characteristic vectors, say the (column) vectors in B . Unfortunately, even though **normalized** characteristic vectors are generally unique subject to normalization, those corresponding to zero roots are not unique! If B_r , a matrix of dimension $q \times r$ is such a matrix, then so is $B_r Q$, where Q is any nonsingular matrix. Thus, it may be argued that what one finds by such a procedure is **only a basis** of the null space of $S_{0,k} \Sigma S'_{0,k}$.¹³ To find the maximal number of zero characteristic roots¹⁴ proceed as follows: extract characteristic roots in decreasing order of magnitude, testing for zero, and stop at the first acceptance. The characteristic vectors corresponding to this root as well as the remaining roots constitute the (maximal rank) estimator of the matrix of cointegrating vectors.

The particular estimator proposed for $S_{0,k} \Sigma S'_{0,k}$ is

$$M_T = \frac{1}{T} \sum_{t=1}^T \left(\frac{X'_t X_t}{t} \right) \quad (55)$$

whose expectation is

$$EM_T = S_{0,k} \Sigma S_{0,k} + \left(\frac{1}{T} \sum_{t=1}^T \frac{1}{t} \right) D. \quad (56)$$

Since the last term behaves like $\ln T/T$, it follows that M_T is an **asymptotically unbiased** estimator of $S_{0,k} \Sigma S'_{0,k}$. It is further conjectured that the limiting distribution of M_T is

$$Q \Lambda^{(1/2)} \left(\int_0^1 \left[\frac{1}{\tau} B(\tau) B'(\tau) \right] d\tau \right) \Lambda^{(1/2)} Q', \text{ where } \Phi = Q \Lambda Q', \Phi = S_{0,k} \Sigma S'_{0,k},$$

¹³ This feature of cointegrating vectors is also pointed out in Johansen (1991), although the context there is quite different.

¹⁴ Because many canned software is programed to extract roots *seriatim* beginning with the largest root, we state the testing procedures as proceeding from the largest to the smallest roots. If the number of zero roots is rather small it would be preferable to proceed from the smallest to the largest. In this case the test procedure would be: test for zero, beginning with the smallest root and stop at the first rejection of the hypothesis.

and B is a standard multivariate (vector) Brownian motion (SMBM). It is further conjectured that the limiting distribution of

$$M_T^* = \frac{1}{T^2} X' X, \quad X = (X_t), \quad t = 1, 2, \dots, T, \quad (57)$$

whose limiting expectation is only **one half** of $S_{0,k} \Sigma S_{0,k}'$, is given by

$$Q \Lambda^{(1/2)} \left(\int_0^1 [B(\tau) B'(\tau)] d\tau \right) \Lambda^{(1/2)} Q', \quad \text{where } \Phi = Q \Lambda Q', \quad \Phi = S_{0,k} \Sigma S_{0,k}',$$

The reason for preferring, at this stage, M_T to M_T^* is that the former, on the average gives a “cleaner” estimator of Φ than does the latter, owing to the fact that

$$EM_T = \Phi + D \left(\frac{1}{T} \sum_{t=1}^T \frac{1}{t} \right), \quad EM_T^* = \left(\frac{T^2 + T}{2T^2} \right) \Phi + \frac{1}{T} D. \quad (58)$$

Oddly enough the preceding suggests an informal unit root test, that ties the current discussions of cointegration to the phenomena discussed in R. A. Stone (1947). To facilitate this task, let

$$X_t^* = \frac{1}{\sqrt{t}} X_t, \quad X^* = (X_t^*), \quad t = 1, 2, 3, \dots, T,$$

and note that if we define M_T **in terms of** X^* , the cointegration procedure involves the extraction of, and tests on, the smallest roots. If we define M_T **in terms of** X , we have the procedure Stone employed for determining the number of “significant” principal components or, conversely, the extent of the near singularity of the covariance matrix of the vector X_t , when the latter is assumed to be an $I(0)$ process!

3.1 Cointegration and Stone’s Procedure

To distinguish the two procedures let

$$X^* = N X, \quad N = \text{diag}(t^{-(1/2)}), \quad t = 1, 2, 3, \dots, T \quad (59)$$

and define

$$M_{T,(0)} = \frac{1}{T} X' X, \quad M_{T,(1)} = \frac{1}{T} X' N^2 X. \quad (60)$$

Stone’s procedure, dressed up in contemporary clothing, entails the determination of the dimension of the null space of the expectation of $M_{T,(0)}$, **on the assumption that** X_t is an $I(0)$ process; cointegration, on the

other hand involves the determination of the dimension of the null space of $M_{T,(1)}$, **on the assumption that** X_t is an $I(1)$ process.

It is clear that the two procedures are connected and, moreover, it is intuitively clear that if, following Stone,¹⁵ we establish near collinearity, we would also expect to establish near cointegration. Thus apart from the formal tests that may be carried out, as a practical matter, once near collinearity is established there is no point in proceeding with cointegration tests!

We formalize this discussion below.

Proposition 3. Consider the matrices $M_{T,(i)}$, $i = 0, 1$, of the previous discussion; the matrix $M_{T,(0)} - M_{T,(1)}$, conditionally on the sample, is positive semidefinite.

Proof: Neglecting the factor $(1/T)$ we find

$$X'X - X'N^2X = X'PX, \quad P = \text{diag}\left(\frac{t-1}{t}\right), \quad t = 1, 2, \dots, T.$$

Since $P \geq 0$, it follows that $X'PX \geq 0$.

q.e.d.

Corollary 2. If, conditionally on the sample, $M_{T,(0)}$ has r_0 roots that obey $\lambda_j^{(0)} \leq \delta$ then there exist $r_1 \geq r_0$ characteristic roots of $M_{T,(1)}$, say $\lambda_s^{(1)}$ such that $\lambda_s^{(1)} \leq \delta$.

Proof: We note, Bellman (1960) p. 115, that since

$$X'X = X'N^2X + X'PX, \tag{61}$$

and both matrices on the right are at least positive semidefinite, it follows that

$$\lambda_j^0 \geq \lambda_j^{(1)}, \quad j = 1, 2, 3, \dots, q. \tag{62}$$

It follows, therefore, that

$$\lambda_j^{(1)} \leq \lambda_j^{(0)} \leq \delta, \quad j = 1, 2, 3, \dots, r_0. \tag{63}$$

¹⁵ Stone examined 17 US macro series and concluded that the(ir) first principal component "explained" about .8076 of the trace of the (sample) covariance matrix of these series; the second component .1059 and the third .0609, or together, they accounted for .9744 of the trace of their (sample) covariance matrix, or the sum of their sample variances. If we are satisfied that the remaining roots are insignificantly different from zero, the null space of the covariance matrix of the 17 random variables in question in 13! See, for example the citation on Stone, or Dhrymes (1970), p. 64.

q.e.d.

Remark 8. Evidently, the implications to be derived from Corollary 2, above are suggestive, but not conclusive. More work is required to establish the limiting distribution of the roots of $M_{T,(0)}$ and $M_{T,(1)}$ and the connection between them, noting the different null hypotheses that characterize such tests. Such issues as well as comparisons of the procedure suggested above relative to those currently in the literature for cointegration tests, will be explored at a subsequent paper.

References

- Anderson, T. W. (1951), "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions", *Annals of Mathematical Statistics*, vol. 22, pp. 327-351.
- (1976), "Estimation of Linear Functional Relationships", *Journal of the Royal Statistical Society, Series B*, vol. 38, pp. 1-36.
- Bellman, R. (1960), *Introduction to Matrix Analysis*, New York: McGraw Hill.
- Dhrymes, P. J. (1970), *Econometrics: Statistical Foundations and Applications*, New York: Harper and Row, corrected edition (1974) New York: Springer Verlag.
- (1971), *Distributed Lags: Problems of Estimation and Formulation* (1971), San Francisco: Holden-Day; second edition, Amsterdam: North Holland (1981).
- (1978), *Mathematics for Econometrics*, New York: Springer Verlag; second edition (1984).
- (1994), "Topics in Time Series", mimeo.
- Dickey, D. A., D.W. Jensen and D. Thornton (1991), "A Primer on Cointegration with an Application to the Demand for Money", *Federal Reserve Bank of St. Louis Review*, vol. 73, pp. 58-78.
- Engle, R. F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, vol. 55, pp. 251-276.
- Izenman, A. J. (1975), "Reduced Rank Regression for the Multivariate Linear Model", *Journal of Multivariate Analysis*, vol. 5, pp. 248-264.
- Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, vol. 12, pp. 231-254.
- (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, vol. 59, pp. 1551-1580.
- Johansen, S. and K. Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money", *Oxford Bulletin of Economics and Statistics*, pp. 169-210.

- Perman, R. (1991), "Cointegration: An Introduction to the Literature", *Journal of Economic Studies*, vol. 18, pp. 3-30.
- Stock, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, vol. 55, pp. 1035-1056.
- Stock, J. H. and M.W. Watson (1988), "Testing for Common Trends", *Journal of the American Statistical Association*, vol. 89, pp. 1097-1107.
- Stone, R. A. (1947), "On the Interdependence of Blocks of Transactions", *Journal of the Royal Statistical Society*, supplement, vol. 9, pp. 1-32.
- Tso, M.K.S. (1981), "Reduced-rank Regression and Canonical Analysis", *Journal of the Royal Statistical Society, Series B*, vol. 43, pp. 183-189.

1994-1995 Discussion Paper Series

Department of Economics
Columbia University
1022 International Affairs Bldg.
420 West 118th Street
New York, N.Y., 10027

The following papers are published in the 1994-95 Columbia University Discussion Paper series which runs from early November to October 31 (Academic Year). Domestic orders for discussion papers are available for purchase at \$8.00 (US) each and \$140.00 (US) for the series. Foreign orders cost \$10.00 (US) for individual paper and \$185.00 for the series. To order discussion papers, please send your check or money order payable to Department of Economics, Columbia University to the above address. Be sure to include the series number for the paper when you place an order.

- 708. Trade and Wages: Choosing among Alternative Explanations
Jagdish Bhagwati
- 709. Dynamics of Canadian Welfare Participation
Garrey F. Barret, Michael I. Cragg
- 710. Much Ado About Nothing? Capital Market Reaction to Changes in
Antitrust Precedent concerning Exclusive Territories.
Sherry A. Glied, Randall S. Kroszner
- 711. The Cost of Diabetes
Matthew Kahn
- 712. Evidence on Unobserved Polluter Abatement Effort
Matthew E. Kahn
- 713. The Premium for Skills: Evidence from Mexico
Michael Cragg
- 714. Measuring the Incentive to be Homeless
Michael Cragg, Mario Epelaum
- 715. The WTO: What Next?
Jagdish Bhagwati
- 716. Do Converters Facilitate the Transition to a New Incompatible Technology?
A Dynamic Analysis of Converters
Jay Phil Choi
- 716A. Shock Therapy and After: Prospects for Russian Reform
Padma Desai
- 717. Wealth Effects, Distribution and The Theory of Organization
-Andrew F. Newman and Patrick Legros

1994-95 Discussion Paper Series

- 718. Trade and the Environment: Does Environmental Diversity Detract from the Case for Free Trade?
-Jagdish Bhagwati and T.N. Srinivasan (Yale Univ)
- 719. US Trade Policy: Successes and Failures
-Jagdish Bhagwati
- 720. Distribution of the Disinflation of Prices in 1990-91 Compared with Previous Business Cycles
-Philip Cagan
- 721. Consequences of Discretion in the Formation of Commodities Policy
-John McLaren
- 722. The Provision of (Two-Way) Converters in the Transition Process to a New Incompatible Technology
-Jay Pil Choi
- 723. Globalization, Sovereignty and Democracy
-Jagdish Bhagwati
- 724. Preemptive R&D, Rent Dissipation and the "Leverage Theory"
-Jay Pil Choi
- 725. The WTO's Agenda: Environment and Labour Standards, Competition Policy and the Question of Regionalism
-Jagdish Bhagwati
- 726. US Trade Policy: The Infatuation with FTAs
-Jagdish Bhagwati
- 727. Democracy and Development: New Thinking on an Old Question
-Jagdish Bhagwati
- 728. The AIDS Epidemic and Economic Policy Analysis
-David E. Bloom, Ajay S. Mahal
- 729. Economics of the Generation and Management of Municipal Solid Waste
-David E. Bloom, David N. Beede
- 730. Does the AIDS Epidemic Really Threaten Economic Growth?
-David E. Bloom, Ajay S. Mahal
- 731. Big-City Governments
-Brendan O'Flaherty
- 732. International Public Opinion on the Environment
-David Bloom

1994-95 Discussion Paper Series

- 733. Is an Integrated Regional Labor Market Emerging in the East and Southeast Asia?
-David Bloom, Waseem Noor
- 734. Migration, Integration and Development
-Abhijit V. Banerjee, Andrew Newman
- 735. Infrastructure, Human Capital and International Trade
-Ronald Findlay
- 736. Short Ballots: Why Mayors Are in Charge of So Many Different Things
-Brendan O'Flaherty
- 737. Demand for Environmental Goods: Evidence from Voting Patterns on California Initiatives
-Matthew Kahn and John Matsusaka
- 738. Inflation and Stabilization in Poland 1990 - 1995
-S. Wellisz
- 739. Particulate Pollution Trends in the 1980's
-M. Kahn
- 740. Why has Wage Dispersion Grown in Mexico? Is it the Incidence of Reforms or the Growing Demand for Skills?
-M.I. Cragg and M. Epelbaum
- 741. Russia's Transition Toward the World Economy:
-P. Desai
- 742. Poland: Transition and Integration in the World Economy
-S. Wellisz
- 743. Team Effects on Compensation: An Application to Salary Determination in the National Hockey League
-T. Idson and L.H. Kahane
- 744. Economic Development, Growth of Human Capital, and the Dynamics of the Wage Structure
-J. Mincer
- 745. Technology Transfer with Moral Hazard
-J.P. Choi (Oct.)
- 746. Risk Preference Smoothness and the APT
-L.F. Herk, Columbia University, Institute on East Central Europe
- 747. A Characterization of Cointegration
P. Dhrymes (Oct.)